

TD6

Enoncé du TD6

Les données se trouvent dans le fichier « tab.csv ».

On dispose de données départementales. Pour chaque département métropolitain, on connaît les votes des premiers tours des élections présidentielles en 2017 (d'après les résultats officiels fournis par le Ministère de l'Intérieur).

1. Quels sont les objectifs d'une AFC sur un tel tableau de données ?

L'AFC permet d'étudier les départements qui se ressemblent et ceux qui sont les plus distincts les uns des autres en fonction des résultats électoraux. Elle permet aussi d'étudier les candidats à l'élection présidentielle, en identifiant celles et ceux qui ont des résultats similaires et distincts dans les différents départements : les candidats qui se ressemblent ont des résultats relativement plus élevés (resp. plus faibles) dans les mêmes départements (attention, on parle ici relativement).

2. Mettre en oeuvre une AFC

Comme d'habitude, on commence par charger les packages R nécessaires à l'analyse :

```
library("FactoMineR")
library("explor")
library(ggplot2)
library(readr)
library(factoextra)
library(dplyr)
library(tibble)
```

On peut ensuite indiquer le répertoire où sont stockées les données et charger les données. On utilise la fonction `column_to_rownames` pour dire à R que la première colonne (nom des départements) doit être considérée comme le nom des lignes du tableau.

```
# Attribution du répertoire de travail : c'est l'emplacement physique où se trouvent les données
setwd("E:/Enseignements/Cours UVSQ/Année 2023-2024/L3 Sociologie Analyse des données/TD/TD1")

# AFC sur les résultats départementaux du premier tour des élections présidentielles
# Lecture du fichier csv.
tab<-read_csv(file="elections.csv", locale=locale(encoding="latin1"))
tab<-column_to_rownames(tab, var="Départements")
```

On lance l'AFC à l'aide de la fonction `CA` du package `FactoMineR`. Attention, il y a une colonne "numéro du département" qui est présente dans ce tableau. On souhaite la conserver pour la suite, mais on n'en tient pas compte ici.

```
# Analyse factorielle des correspondances
res.afc<-CA(tab[,-1], graph=F)
```

Pour analyser l'AFC, on peut utiliser les fonctions `fviz` ou `explor`.

3. Combien d'axes retenir ?

Pour le critère de Kaiser, on regarde les valeurs propres qui ont une valeur supérieure à la valeur propre moyenne :

```
vp<-res.afc$eig # On regarde lesquelles sont supérieures à l'inertie moyenne
vp[,1]>mean(vp[,1])
```

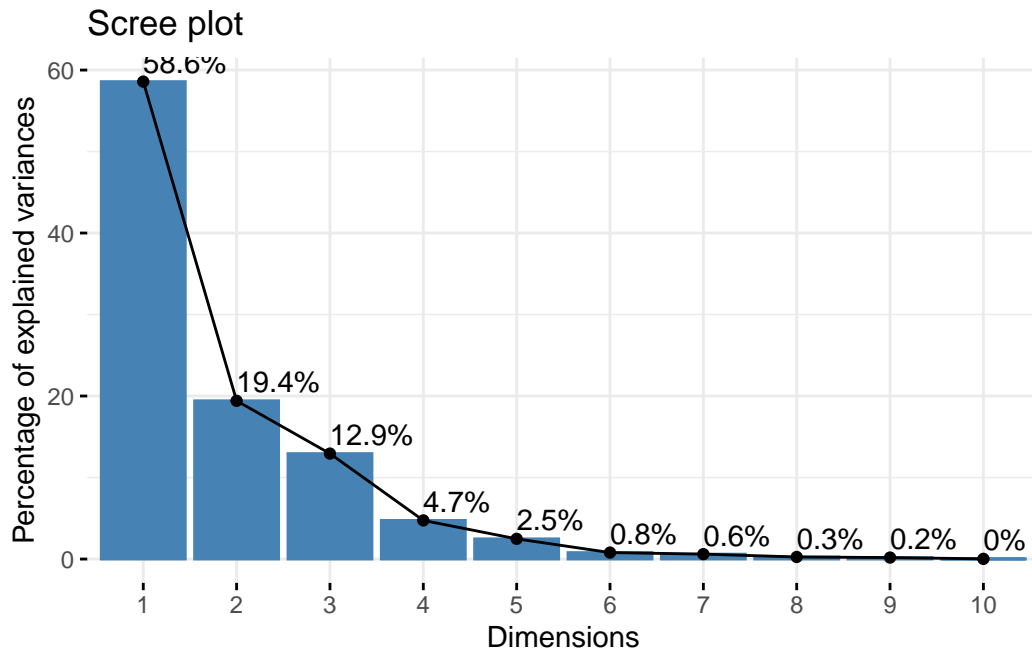
dim 1	dim 2	dim 3	dim 4	dim 5	dim 6	dim 7	dim 8	dim 9	dim 10
TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

Cela nous pousserait à examiner 3 axes.

Pour le critère du coude, on peut regarder l'histogramme des valeurs propres :

```
##Le scree plot des valeurs propres

fviz_eig(res.afc, addlabels=T, ncp =10)
```



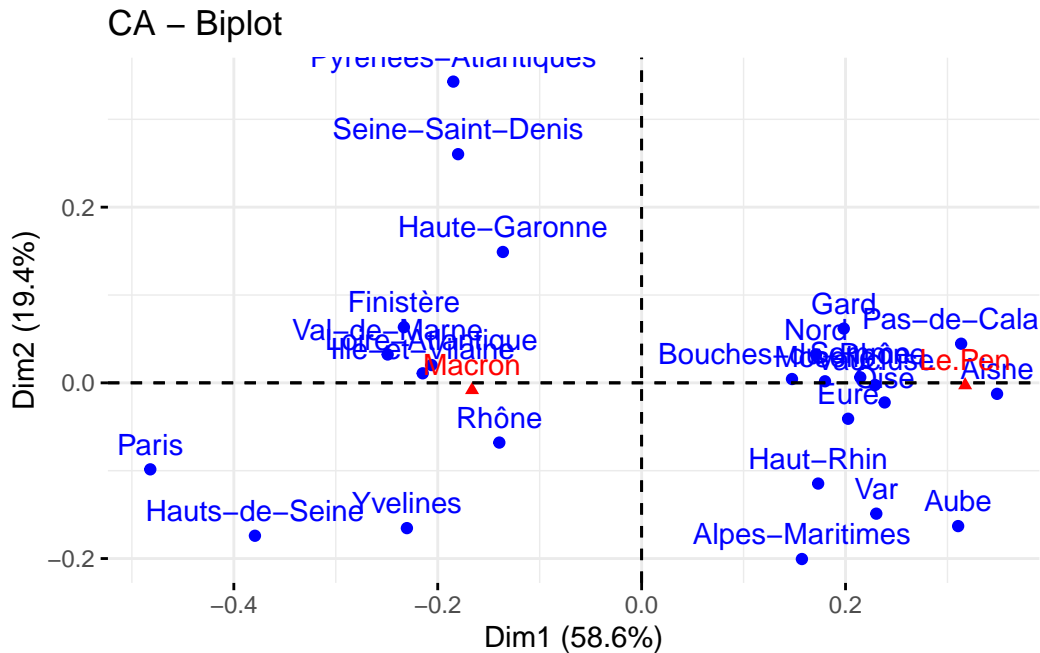
On a un premier décrochage entre la première et seconde dimension de l'AFC : la première est prépondérante. Ensuite un décrochage entre la troisième et la quatrième : il semble que trois axes aient une valeur propre importante.

4. Interprétation des axes

Par simplicité, regardons ici les deux premiers axes.

Quelles sont les lignes (départements) et les colonnes (candidats) qui ont les plus fortes contributions sur la première dimension (contribution supérieure à la contribution moyenne) ?

```
##Visualiser la contribution des lignes et colonnes qui ont une contribution supérieure à
rowcontrib1<-rownames(res.afc$row$contrib)[res.afc$row$contrib[,1]>100/nrow(tab)]
colcontrib1<-rownames(res.afc$col$contrib)[res.afc$col$contrib[,1]>100/ncol(tab)]
fviz_ca_biplot(res.afc,axes=c(1,2),
               select.row = list(name = rowcontrib1),
               select.col=list(name=colcontrib1))
```



Cet axe oppose les départements qui ont relativement davantage voté pour Macron au premier tour versus les départements qui ont davantage voté Le Pen au premier tour. Parmi les départements qui ont davantage voté Macron par rapport aux autres, ceux de l’Île de France et de Bretagne notamment. Parmi les départements qui ont davantage voté Le Pen, ceux du Nord de la France et du sud-est.

Le principal clivage électoral à l’échelle départementale en 2017 n’est donc pas un clivage droite-gauche, mais un clivage entre le président sortant (centriste/libéral...) et l’extrême-droite.

Mais comment les candidats de gauche et de droite se positionnent sur cet axe ? Même s’ils ont une contribution plus faible à la structuration de l’axe, on peut les projeter pour voir où ils se positionnent par rapport aux différents départements. On peut utiliser l’indicateur du \cos^2 (valeur entre 0 et 1) comme qualité de représentation pour voir s’ils sont plus ou moins bien représentés sur cet axe.

A part Macron et Le Pen, Hamon est plutôt bien représenté sur cet axe, et viennent ensuite Dupont-Aignan, Arthaud et Fillon, mais qui sont nettement moins bien représentés. Les autres sont définitivement mal représentés sur cet axe (leur coordonnée est ainsi proche de l’origine).

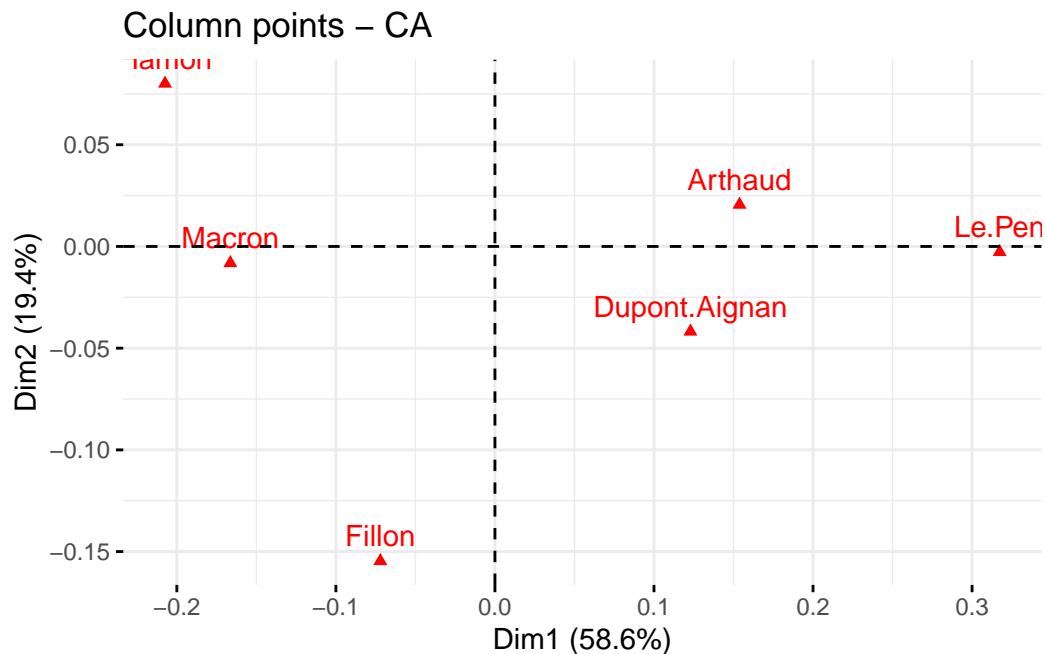
```
data.frame(Axe1_cos2=sort(res.afc$col$cos2[,1],decreasing = T))
```

Axe1_cos2

Le.Pen	0.9924666859
Macron	0.9304129092
Hamon	0.7199111798
Dupont.Aignan	0.3054563501
Arthaud	0.2407110932
Fillon	0.1490396460
Mélenchon	0.0674743898
Cheminade	0.0662298705
Poutou	0.0110155375
Lassalle	0.0009079383
Asselineau	0.0002071592

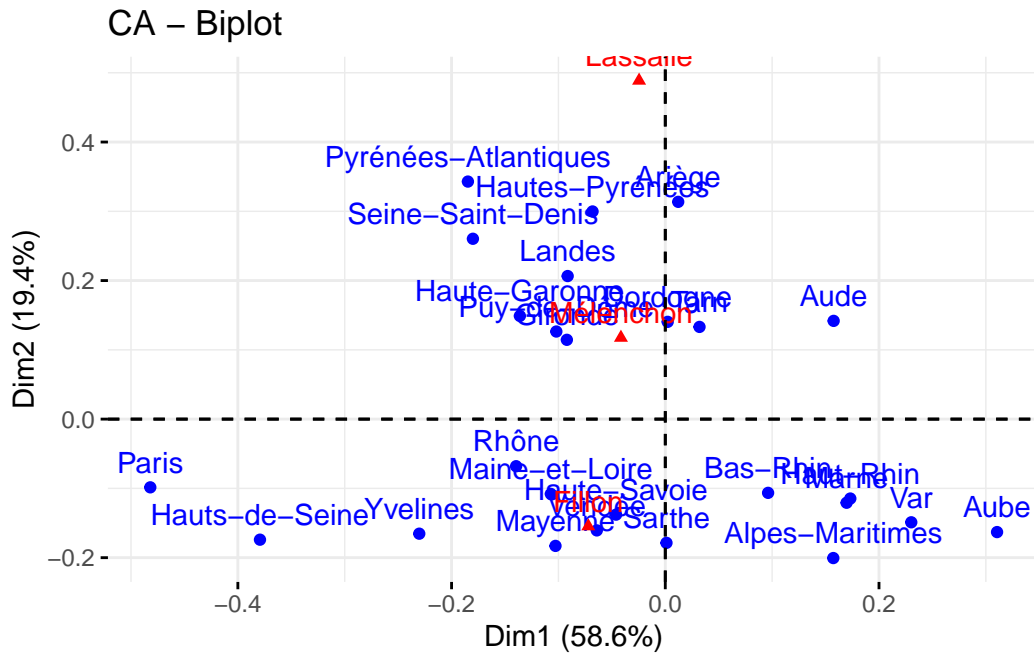
```
colcos1<-rownames(res.afc$col$cos2)[res.afc$col$cos2[,1]>.1]

fviz_ca_col(res.afc,axes=c(1,2),select.col=list(name=colcos1))
```



Hamon fait des scores relativement plus élevés là où Macron fait des scores élevés. C'est aussi le cas de Fillon, même si ce n'est pas sa seule particularité. En revanche, Arthaud et Dupont-Aignan font des scores relativement plus élevés là où Le Pen fait des scores élevés, même si attention leur qualité de représentation sur l'axe est plus faible...

```
##Visualiser la contribution des lignes et colonnes qui ont une contribution supérieure à
rowcontrib2<-rownames(res.afc$row$contrib)[res.afc$row$contrib[,2]>100/nrow(tab)]
colcontrib2<-rownames(res.afc$col$contrib)[res.afc$col$contrib[,2]>100/ncol(tab)]
fviz_ca_biplot(res.afc,axes=c(1,2),
               select.row = list(name = rowcontrib2),
               select.col=list(name=colcontrib2))
```



L'axe 2 révèle un axe lié à l'ancrage géographique des candidats : Lassalle est associé au haut de l'axe 2, Fillon au bas de cet axe. Mélenchon est dans une position intermédiaire en haut de l'axe 2.

```
data.frame(Axe2_cos2=sort(res.afc$col$cos2[,2],decreasing = T))
```

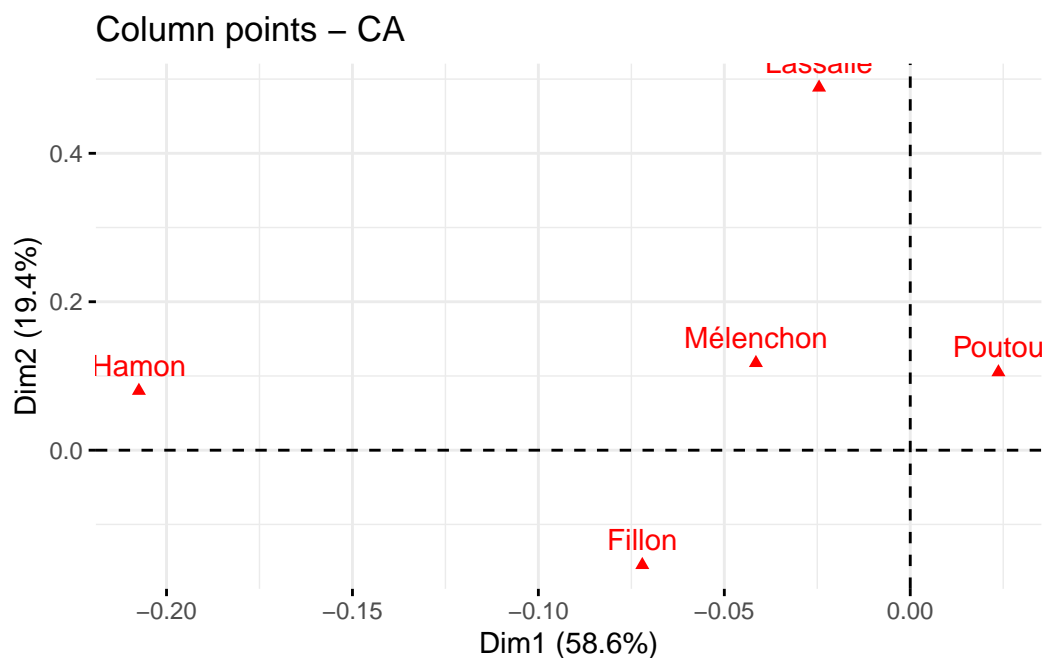
	Axe2_cos2
Fillon	6.864982e-01
Mélenchon	5.409861e-01
Lassalle	3.605284e-01
Poutou	2.164280e-01
Hamon	1.072459e-01
Dupont.Aignan	3.534046e-02
Arthaud	4.296670e-03

Asselineau	2.723210e-03
Macron	2.242355e-03
Cheminade	1.165384e-03
Le.Pen	7.262012e-05

En plus de ces candidats qui contribuent beaucoup à l'axe, on voit que Poutou et Hamon ont aussi une plutôt bonne qualité de représentation...

```
colcos2<-rownames(res.afc$col$cos2)[res.afc$col$cos2[,2]>.1]

fviz_ca_col(res.afc,axes=c(1,2),select.col=list(name=colcos2))
```

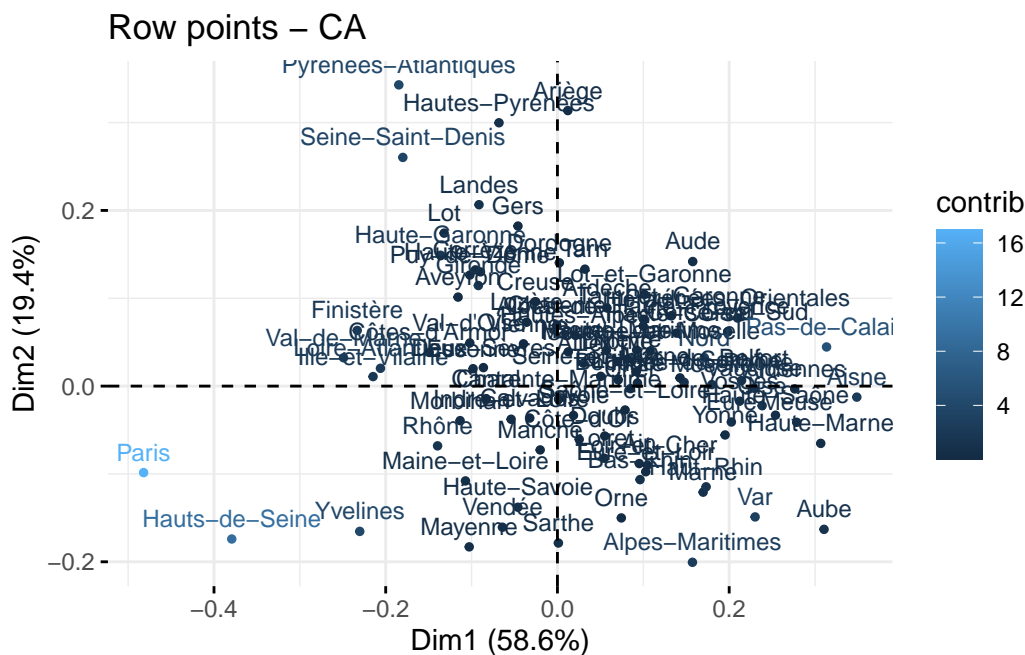


Même s'ils sont positionnés de manière assez proches de l'origine sur l'axe 2, on voit que ces candidats partagent une certaine similarité dans leurs scores électoraux avec Lassalle. Pour ce dernier, sa spécificité ressort de son ancrage électoral dans les Pyrénées dont il est originaire, tandis que Fillon avait son "fief électoral dans la Sarthe, et qu'on voit bien qu'il est associé sur l'axe 2 à des départements qui sont traditionnellement plutôt de "droite" (la Vendée, les Yvelines, les Hauts de Seine...). Derrière un axe d'ancrage géographique des candidats se cache donc aussi un clivage droite-gauche sur ce second axe !

5. Quel département choisit-on si on désire faire un sondage préélectoral de la meilleure qualité possible dans un département « représentatif » ?

Un département représentatif serait un département qui se différencie peu de l'origine sur les axes de l'analyse factorielle. Proche de l'origine, c'est un département "moyen" où les électeurs ne votent ni plus ni moins que la moyenne nationale pour les différents candidats qui ressortent sur les axes.

```
fviz_ca_row(res.afc,axes=c(1,2),col.row="contrib",pointsize=1,labels=3)
```



```
#La coloration des départements en fonction de la contribution correspond à la contribution
```

La figure n'est pas aisée à lire mais c'est la Charente-Maritime qui est le plus proche de l'origine sur le premier et deuxième axe. On pourra utiliser la fonction `explor()` pour repérer plus facilement les points sur le plan factoriel.

Autre méthode plus robuste, car ici on a seulement regardé graphiquement le département moyen sur le premier plan factoriel. Un profil-moyen correspond à un profil qui contribue peu globalement à la dispersion des points du nuage factoriel (sur tous les axes), relativement à son poids démographique (nombre d'électeur total dans le département). En effet, on sait que dans une AFC la contribution des lignes et des colonnes dépend de leur poids (un plus grand poids correspond à une plus grande contribution, c'est d'ailleurs ce qui explique que Macron

ait une contribution plus importante sur l'axe 1 que Hamon par exemple, le premier a plus d'électeurs). Donc on va chercher le département qui a la contribution à l'inertie globale du nuage des départements la plus faible, relativement à son poids dans le total des électeurs.

```
## Poids de chaque département
poids<-res.afc$call$marge.row
## Contribution à l'inertie de chaque département
contribution<-res.afc$row$inertia/sum(res.afc$row$inertia)
## Ratio contribution poids
ratio<-contribution/poids
## On affiche directement la valeur minimale du ratio
which.min(ratio)
```

Charente-Maritime
17

```
##Si on n'avait pas tenu compte du poids :
names(contribution)<-row.names(tab)
which.min(contribution)
```

Allier
3

On voit que relativement à son poids en nombre d'électeurs, c'est bien la Charente-Maritime qui contribue le moins à l'inertie du nuage, c'est ce département qui a le profil le plus "moyen". Si on n'avait pas tenu compte du poids du nombre total d'électeurs, ça aurait été l'Allier, mais qui contribue peu à l'inertie du fait du faible nombre d'électeurs dans ce département...

5. A l'aide du script R fourni, représenter sous forme cartographique les coordonnées des départements sur les axes retenus.

On va utiliser des fichiers complémentaires, un fond de carte. C'est un type de données particulier qui associe des objets à des coordonnées (géodésiques) pour pouvoir représenter ces objets dans un espace. On pourra par exemple retrouver un référencement de fonds de cartes [ici](#).

```
###Représentation spatiale des résultats

#On charge une base avec les coordonnées spatiales des départements (format shp)
library(sf)
```

```
##Pour zoomer sur l'ile de France
library(ggmapinset)
```

```
setwd("E:/Enseignements/Cours UVSQ/Année 2023-2024/L3 Sociologie Analyse des données/TD/TD6\Cartographie")
dep <- st_read("DEPARTEMENT.shp")
```

Reading layer `DEPARTEMENT' from data source

```
`E:\Enseignements\Cours UVSQ\Année 2023-2024\L3 Sociologie Analyse des données\TD\TD6\Cartographie'
using driver `ESRI Shapefile'
```

Simple feature collection with 96 features and 11 fields

Geometry type: MULTIPOLYGON

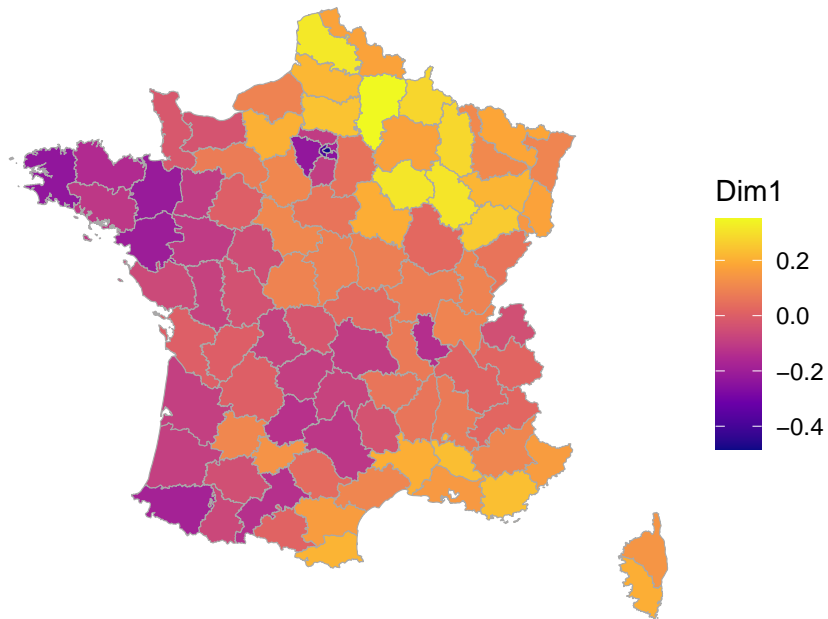
Dimension: XY

Bounding box: xmin: 99217.1 ymin: 6049646 xmax: 1242417 ymax: 7110480

Projected CRS: RGF93 v1 / Lambert-93

```
#On crée une base avec les coordonnées sur les trois premiers axes. On indique l'id du dép
coordafc<-as.data.frame(cbind(tab$Code,res.afc$row$coord[,1:3]))
names(coordafc)<-c("Code","Dim1","Dim2","Dim3")
coordafc<-coordafc %>% mutate_at(vars(Dim1:Dim3),as.numeric)
#On ajoute des 0 pour les codes de département qui n'ont qu'un chiffre pour matcher avec l'axe Dim1
coordafc$Code<-ifelse(nchar(coordafc$Code)==1,paste0("0",coordafc$Code),coordafc$Code)
##On fait une jointure avec la base du fond de carte
dep2 <- dep %>% left_join(coordafc,by=c("CODE_DEPT"="Code"))
```

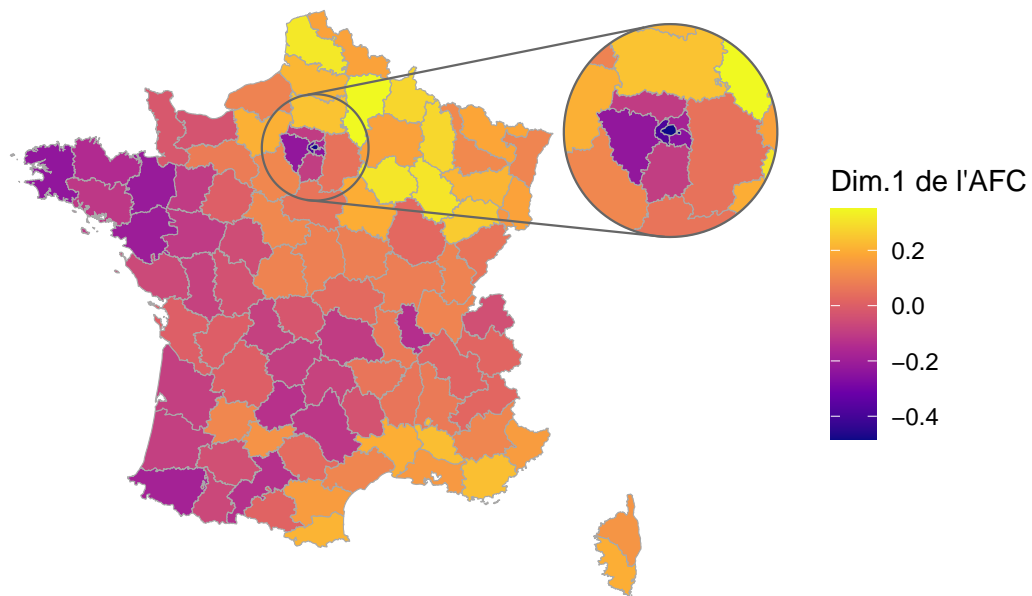
```
d1 <- ggplot(dep2) +
  geom_sf(aes(fill=Dim1),colour="darkgrey")+
  scale_fill_viridis_c(option="C")+
  theme_void()
d1
```



Cette première carte est intéressante, mais on peut vouloir zoomer sur l'Ile de France.

```
# Pour faire le zoom autour de Paris sur la carte:
inset_centre <- sf::st_centroid(sf::st_geometry(dep2)[dep2$NOM_DEPT == "PARIS"])

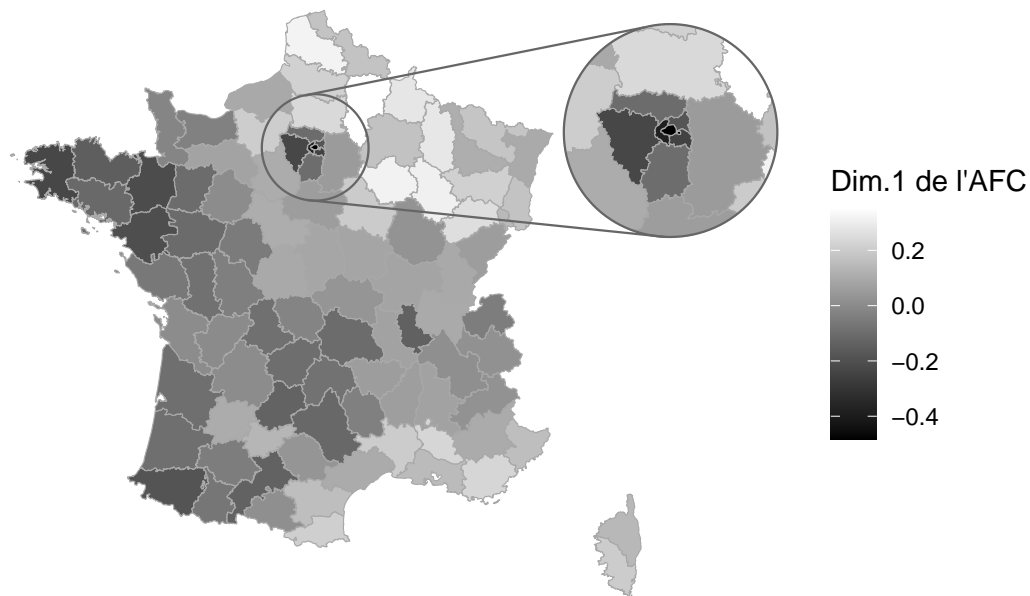
#La même carte avec le zoom autour de Paris:
d1<-ggplot(dep2) +
  geom_sf_inset(aes(fill = Dim1),colour="darkgrey") +
  geom_inset_frame() +
  coord_sf_inset(inset = configure_inset(centre = inset_centre, scale = 2, units = "mi",
                                         translation = c(400, 0), radius = 60))+
  scale_fill_viridis_c(option="C")+
  labs(fill="Dim.1 de l'AFC")+
  theme_void()
d1
```



C'est mieux ! On repère ainsi les clivages électoraux résumés sur le premier axe de l'AFC.

On peut préférer une représentation en noir et blanc :

```
#La même carte en noir et gris
d1<-ggplot(dep2) +
  geom_sf_inset(aes(fill = Dim1),colour="darkgrey") +
  geom_inset_frame() +
  coord_sf_inset(inset = configure_inset(centre = inset_centre, scale = 2, units = "mi",
                                         translation = c(400, 0), radius = 60))+
  scale_fill_gradient(low="black",high="white")+
  labs(fill="Dim.1 de l'AFC")+
  theme_void()
d1
```



On a ici représenté les coordonnées du premier axe de l'AFC sur la carte. La tonalité de l'échelle des couleurs correspond ainsi à une nuance de position entre le département le plus à gauche sur l'axe 1 et le plus à droite sur l'axe 2. On aurait pu discrétiser les coordonnées et avoir ainsi des classes de département par exemple.

Enfin, on peut vouloir vérifier la pertinence de l'AFC en projetant les résultats électoraux bruts. Par exemple, pour le score de Macron :

```
#On calcule d'abord le pourcentage des votes en faveur de chaque candidat par departement
##On fait une jointure avec la base du fond de carte

tabprop<-as.data.frame(cbind(Code=tab$Code,tab[-1]/rowSums(tab[,-1])*100))
tabprop$Code<-ifelse(nchar(tabprop$Code)==1,paste0("0",tabprop$Code),tabprop$Code)

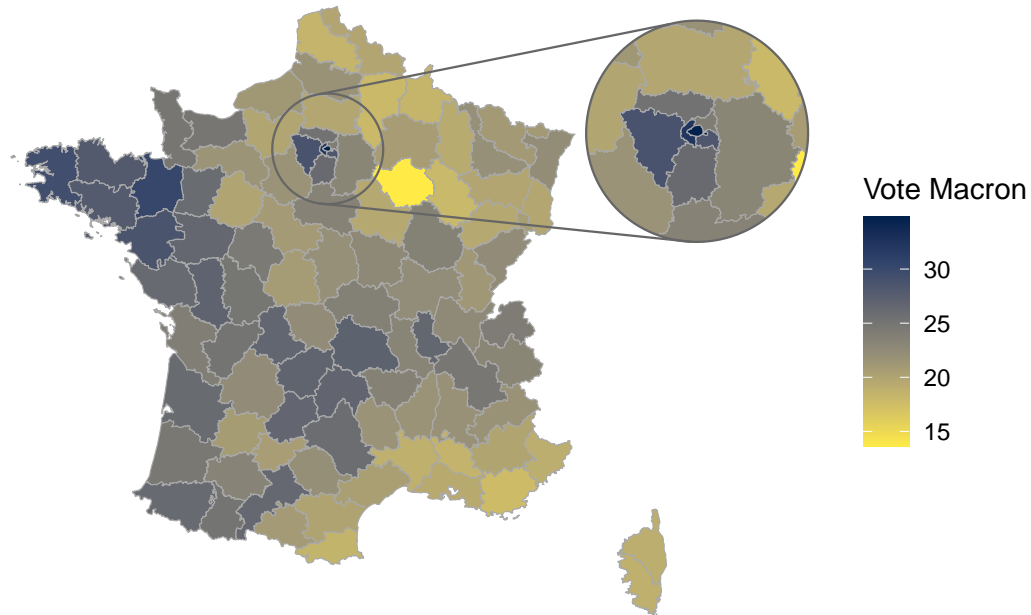
dep3 <- dep %>% left_join(tabprop,by=c("CODE_DEPT"="Code"))

macron<-ggplot(dep3) +
  geom_sf_inset(aes(fill = Macron),colour="darkgrey") +
  geom_inset_frame() +
  coord_sf_inset(inset = configure_inset(centre = inset_centre, scale = 2, units = "mi",
                                         translation = c(400, 0), radius = 60))+
  scale_fill_viridis_c(option="E",direction =-1)+
```

```

labs(fill="Vote Macron")+
theme_void()
macron

```

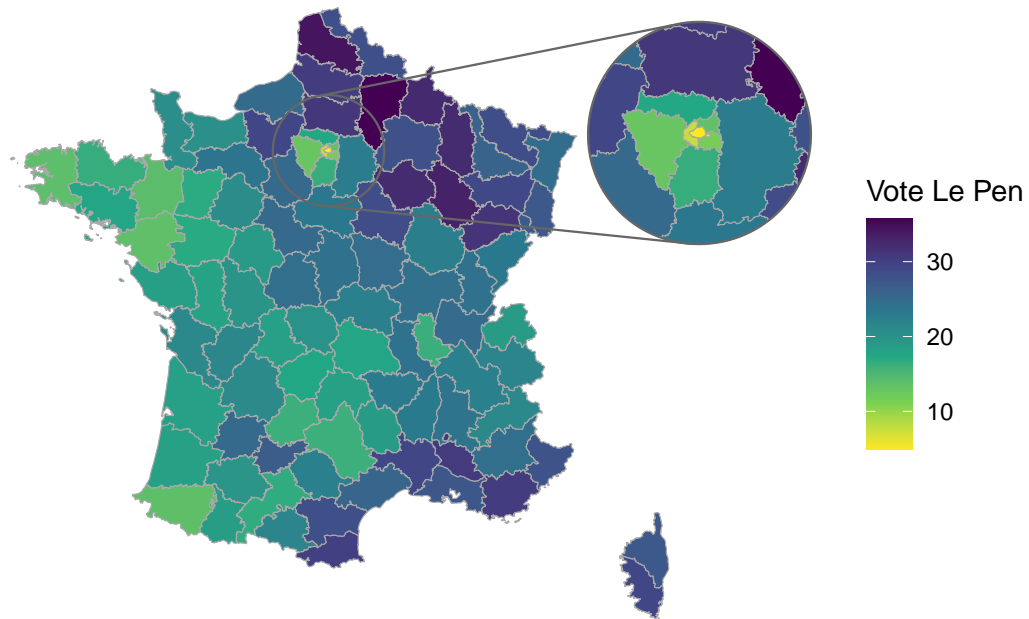


Et de Le Pen, on voit grossièrement que les deux candidats ont des scores élevés dans des département différents.

```

lepen<-ggplot(dep3) +
  geom_sf_inset(aes(fill = Le.Pen),colour="darkgrey") +
  geom_inset_frame() +
  coord_sf_inset(inset = configure_inset(centre = inset_centre, scale = 2, units = "mi",
                                         translation = c(400, 0), radius = 60))+
  scale_fill_viridis_c(direction=-1)+
  labs(fill="Vote Le Pen")+
  theme_void()
lepen

```



Note : on s'est retenu d'interpréter ici les clivages électoraux à l'échelle des départements en fonction des propriétés socio-démographiques de ces mêmes départements qui sont importants. On pourrait par exemple étendre cette analyse en ajoutant des colonnes supplémentaires avec les caractéristiques socio-démographiques supplémentaires. Pour une analyse récente de la sociologie électorale à l'échelle des communes françaises, on pourra se reporter à Cagé, J., & Piketty, T. (2023). Une histoire du conflit politique: élections et inégalités sociales en France, 1789-2022. Paris XIXe: Éditions du Seuil.